

REMARKS

Applicant respectfully requests reconsideration and allowance in view of the foregoing amendments and following remarks. In the Office Action, mailed June 26, 2003, the Examiner rejected claims 1-39. By this amendment, claims 1, 2, 8, 9, 10, 12 and 14 have been amended. Following entry of these amendments, claims 1-39 will be pending in the application.

Claim Rejections under 35 U.S.C. §103(a)

In the Office Action, the Examiner rejected claims 1-39 under 35 U.S.C. §103(a) as allegedly being unpatentable over U.S. Patent No. 5,924,090 to Krellenstein (hereinafter "Krellenstein") in view of the article entitled "Web Document Clustering: A Feasibility Demonstration" to Zamir et al. (hereinafter "Zamir"). Applicants respectfully traverse the rejections of claims 1-39 and note the following standards for a proper §103(a) rejection.

A §103(a), or obviousness, rejection is proper only when "the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which the subject matter pertains." 35 U.S.C. §103(a). The Examiner must make out a *prima facie* case for obviousness. The *en banc* Federal Circuit has held that "structural similarity between claimed and prior art subject matter, proved by combining references or otherwise, where the prior art gives reason or motivation to make the claimed compositions, creates a *prima facie* case of obviousness." *In re Dillon*, 16 U.S.P.Q. 2d 1897, 1901 (CAFC 1990).

Further, the mere fact that references can be combined or modified does not render the resultant combination obvious unless the prior art also suggests the desirability of the combination. *In re Mills*, 916 F.2d 680, 16 U.S.P.Q.2d 1430 (Fed. Cir. 1990). Likewise, if the proposed modification would render the prior art invention being modified unsatisfactory for its intended purpose, then there is no suggestion or motivation to make the proposed modification. *In re Gordon*, 733 F.2d 900, 221 U.S.P.Q. 1125 (Fed. Cir. 1984).

The underlying inquiries into the validity of an obvious rejection are: "(1) the scope and content of the prior art; (2) the level of ordinary skill in the prior art; (3) the differences between the claimed invention and the prior art; and (4) objective evidence of nonobviousness." *In re*

Dembiczak, 175 F.3d 994, 998, (Fed. Cir. 1999).

For at least the reasons stated below and taking into consideration the standards for obviousness presented above, Applicants assert that one of ordinary skill in the art would not have considered Applicants' invention obvious at the time of invention and, therefore, that Applicants' rejected claims 1-39 are not obvious over the prior art of record.

Claim 1

Applicants' amended independent claim 1 recites a method of categorizing an initial collection of documents, each document being represented by a string of characters, that includes the steps of:

- identifying predefined characters in the string of characters from the documents in the initial collection of documents to form identified characters;
- changing the identified characters in the documents in the initial collection of documents to form a preprocessed collection of documents, each of the preprocessed collection of documents represented by a preprocessed string of characters;
- constructing a number of categories from the preprocessed string of characters of the preprocessed collection of documents; and
- assigning each document in the preprocessed collection of documents to a category to form a hierarchy of categories of documents.

In rejecting Applicants' independent claim 1, the Examiner refers to Figure 2 and col. 2, ll. 56-65, of Krellenstein, and p. 3, sect. 3.1, of Zamir. The search method and apparatus disclosed by Krellenstein categorizes records within a database that are pre-classified according to various meta-data attributes (e.g., subject, type, source, and language). In fact, for the Krellenstein categorization to work, each record in the database must be classified per the meta-data attributes (Krellenstein, Abstract, and col. 8, ll. 56-59). Zamir describes a document clustering method it designates as 'suffix tree clustering' that treats the searched documents as a string, making use of proximity information between words (Zamir, p. 1, col. 2, 2d para.). The Zamir suffix tree is constructed using all of the sentences of all of the documents in the collection of documents (Zamir, p. 3, col. 2, 1st para.).

In contrast, claim 1 of the present invention discloses a categorization method that does not require the pre-classification of meta-data attributes to work, as does Krellenstein, and does not use

all sentences of all documents to create a category tree, as does Zamir. Claim 1 of the present invention categorizes the documents based on the string of characters that make up the documents themselves. Applicants' process categorizes one document at a time, to create one category at a time, without first creating the category structure based on all words of all documents. Claim 1 of the present invention performs properly regardless of whether the documents have any meta-data attributes. As previously mentioned, Krellenstein can only categorize documents that have meta-data attributes.

Further, the reason Krellenstein can only categorize documents that have meta-data attributes is because those attributes; or more specifically, the set of all meta-data attribute values of all of the pre-classified documents being categorized, make up the closed set of possible categories into which the documents can be categorized. The Krellenstein categories are predefined into a closed set by the pre-classification process and are not dynamically created during the document categorization process, as in the present invention.

Additionally, the Examiner has failed to make out a *prima facie* case for obvious. The Examiner has not pointed out what motivation there is to combine Krellenstein and Zamir. The likely reason for this is that, combining the teachings of Zamir to Krellenstein would render the Krellenstein invention unsatisfactory for its intended purpose. That is, Krellenstein relies on the pre-classification of the meta-data attributes of each document. These meta-data attributes have pre-defined, specific values. The Krellenstein categorization process would be rendered useless if the Zamir document cleaning process were allowed to act on the pre-classification values. Consequently, Zamir cannot be added to Krellenstein as suggested by the Examiner.

Therefore, for at least the reasons presented above, Applicants request the withdrawal and reconsideration of the claim rejections for independent claim 1. Applicants respectfully submit that independent claim 1 is in a condition for allowance, and respectfully request such a Notice to that effect.

Dependent Claims 2-25

Dependent claims 2-25 all ultimately depend from independent claim 1. The allowability of dependent claim 2-25 thus follows from the allowability of independent claim 1; as such, dependent claims 2-25 are allowable over the art of record.

Specifically, regarding dependent claim 2, the Examiner refers to p. 3, sect. 3.2, of Zamir and Fig. 2 of Krellenstein to support the rejection. First, the Examiner again fails to show the motivation to combine these two references in relation to the limitations of claim 2. Zamir develops a tree of word nodes based on word phrases from all of the documents, simultaneously. Krellenstein uses pre-classified attribute values as pre-determined categories for the documents to be placed into. The Zamir-type tree development and the Krellenstein-type pre-classification category determination are in stark contrast to Applicants' invention. Applicants disclose a systematic process whereby a first document is picked and placed into a temporary category holder. Then, each of the remaining documents are compared and tested against the first document. The remaining documents that test as being similar to the first document are placed into the temporary category with the first document. If a sufficient number of documents end up being in the temporary category, then that temporary category is classified as a 'real' category and 'real' category parameters are determined.

Specifically, regarding dependent claim 4, the Examiner has made a leap of faith that because Zamir discloses changing plural words to singular words, one of ordinary skill in the art would have known to change upper case characters to lower case characters "in order to identify key phrases and enhance user readability." Applicants assert, though, that user readability is reduced by converting capitals to lower case letters, especially after punctuation marks have been removed. Further, many "key phrases" are proper nouns and, therefore, it would not be obvious to remove the capitalization of such phrases to enhance their recognition. Finally, changing capitalized letters to lower case letters defeats one of the express aspects of Krellenstein, which states that "if the capitalization is the same as in the query term, the record ranks higher." (Krellenstein, col. 4, ll 54-55).

Specifically, regarding dependent claim 5, nowhere does Zamir suggest changing non-root words with their root form, especially as these phrases are used in Applicants' specification. The example used in Applicants' specification include changing the word "went" to "go". This example illustrates that converting non-root to root includes converting past tense to present tense. In fact, the example provided in the Zamir teaching uses the word "ate" in each of the three phrases, and does not change this word to "ear", as would be the case in Applicants' invention.

Specifically, regarding dependent claim 6, the Examiner has made a leap of faith that because Zamir discloses changing plural words to singular words, one of ordinary skill in the art would have known to change abbreviations with the spelled-out equivalent "in order to identify key phrases and enhance user readability." Applicants assert, though, that user readability is reduced by converting abbreviations with the spelled-out equivalent, especially for very commonly used abbreviations (e.g., AM/PM after time of day, or BC/AD after calendar year) for which the spelled-out equivalent is not generally recognizable. Further, many "key phrases" include abbreviations and, therefore, it would not be obvious to remove the abbreviations of such phrases to enhance their recognition.

Specifically, regarding dependent claim 7, there is no suggestion or teaching within Zamir and Krellenstein to remove entire words from the document before categorization. Therefore, it would never cross the mind, much less be obvious, of someone skilled in the art to remove entire words.

Specifically, regarding dependent claim 8, and as previously discussed, the Zamir reference creates a tree from phrases of multiple documents and then creates the categories from the tree. This is contrary to Applicants' invention, where one document (the seed document) is chosen to be compared against all remaining documents for similarities, and then used for initializing the category properties.

Specifically, regarding dependent claim 9, Krellenstein uses the pre-classified values of the meta-data attributes as the candidate categories and sorts the documents into these categories. Then Krellenstein weights the candidate categories and determines which ones to display to the user. This is contrary to Applicants' invention, where the categories are dynamically and arbitrarily decided by seed documents and are altered and updated as the categorization process proceeds.

Specifically, regarding dependent claim 11, neither Krellenstein nor Zamir disclose re-categorizing documents in lesser-populated, existing categories after all the documents have been categorized. The categories of Krellenstein are pre-defined by the pre-classification process, and each document, therefore, is limited as to which category it can belong. Thus, once the documents of Krellenstein are placed within its meta-data category, re-shuffling the documents is not relevant. Likewise, in Zamir, once the tree for the documents is established, then the categories are set. Re-

shuffling the documents would defeat the purpose of establishing the tree in the first place.

Specifically, regarding dependent claim 12, Krellenstein uses the pre-classified values of the meta-data attributes as the candidate categories and sorts the documents into these categories. Then Krellenstein weights the candidate categories and determines which ones to display to the user. This is contrary to Applicants' invention, where the categories are dynamically and arbitrarily decided by seed documents and are altered and updated as the categorization process proceeds.

Specifically, regarding dependent claim 14, Krellenstein uses the pre-classified values of the meta-data attributes as the candidate categories and nowhere teaches or suggests using particular strings within the body of the document for this purpose. Further, Krellenstein nowhere suggests using a fractional number for each document type. Zamir uses node names from the constructed tree to set cluster names. In this way, the construction of the tree and how the nodes are interrelated between the words of many documents dictates the cluster names. Thus, using the node names as cluster names is independent of particularly using one string combination of one document as the category property.

Specifically, regarding dependent claim 17, the Examiner points to Krellenstein, col. 6, l. 66, through col. 7, l. 27, for support. However, nowhere in this passage of Krellenstein, or anywhere in Krellenstein, is there a teaching or suggestion that two categories are merged, or promoting sub-categories into a higher level where the higher level does not have enough documents. Krellenstein, by design, groups the documents by similar combinations of meta-data attribute values. It chooses categories containing 20% to 80% of the search results documents for subsequent weighting. The Krellenstein categorization process does not seek out commonality in category property and combine those categories having sufficient similarity. Further, Krellenstein does not promote lower tier categories into upper tier, under-populated, categories. In fact, as shown in Fig. 2 of Krellenstein, subordinate categories are not even created until after the first-cut (i.e., higher level) categories are displayed to the user (step 44) and the user selects a top-level category for display (step 46). Then, and only then, does Krellenstein determine whether there is a sufficient amount of sub-records to warrant a second level of categorization (bottom output of step 50, going to decision block 34). Thus, Krellenstein cannot promote lower tier categories into upper tier categories because the lower tier categories are not created until after an upper tier category is selected.

Specifically, regarding dependent claim 18, Zamir uses the phrases from multiple documents, all at once, to create the tree. There is not a seed document or a first document. In fact, Zamir states, at p. 3, col. 2, ll. 7-9, "in our application, we construct the suffix tree of all the sentences of all the documents in our collection." (emphasis added). Thus, there is not the concept or suggestion of a seed document in Zamir.

Specifically, regarding dependent claim 19, neither Zamir nor Krellenstein teaches or suggests the use of a seed document, much less the means by which the seed document is determined or identified, as in Applicants' claimed invention.

Specifically, regarding dependent claim 20, neither Zamir nor Krellenstein teaches or suggests the use of a temporary category as claimed by Applicants. Applicants' temporary category, as claimed, does not have category properties assigned. It is merely a holding area, or testing area. All categories in Krellenstein are determined by, and have the properties of, the specific pre-classified meta-data values. Likewise, the Zamir nodes, which represent a group of documents and a phrase that is common to all of them, are defined while constructing the tree. Thus, neither Zamir nor Krellenstein contemplate the idea of a temporary category.

Specifically, regarding dependent claims 23-25, Zamir neither teaches or suggests the use of an anchor-text character string as in Applicants' claimed invention. In fact, in Zamir, at page 3, sect. 3.1, the initial document cleaning step strips the "non-word tokens (such as numbers, HTML tags and most punctuation)." Therefore, it would be impossible for Zamir to use an HTML anchor-text string that had previously been stripped from the documents.

Therefore, for at least these reasons presented above, Applicants respectfully submit that dependent claims 2-25 are in a condition for allowance, and respectfully request such a Notice to that effect.

Claims 26-39

In the office action, the Examiner rejected claims 26-39 "on grounds corresponding to the reasons given above for claims 1-25." Therefore, Applicants contend that, for at least all of the reasons for allowability presented above in relation to the rejections of claims 1-25, the art of record neither discloses nor suggests the subject matter of claims 26-39; thus, these claims are allowable over the art of record.


Therefore, for at least these reasons, Applicants respectfully submit that claims 26-39 are in a condition for allowance, and respectfully request such a Notice to that effect.

Conclusion

All objections and rejections having been addressed, it is respectfully submitted that the present application is in a condition of allowance and a Notice to that effect is earnestly solicited. If any points remain in issue which the Examiner feels may be best resolved through a personal or telephone interview, the Examiner is kindly requested to contact the undersigned at the telephone number listed below.

CHARGE STATEMENT: The Commissioner is hereby authorized to charge fees that may be required relative to this application, or credit any overpayment, to our Account 03-3975, Order No. 053684-0300105 (LS-002).

Respectfully submitted,
PILLSBURY WINTHROP LLP

By: 
Ross L. Franks, Reg. No. 47,233
For: David A. Jakopin, Reg. No. 32,995

2550 Hanover Street
Palo Alto, CA 94304-1115
Tel. No.: (650) 233-4897
Fax No.: (650) 233-4545

60337151